

Search Engine Spiders

Search engines need a way to examine the websites that people submit to them. Some search engines like Yahoo!™ employ people to check each website registered with them, but these are in the minority. Most other search engines use specialized software called "spiders" or "robots" to do this. Spiders help search engines deliver accurate search results by determining how relevant a website is to the phrases and keywords a web surfer uses. Spiders "crawl" through websites, analyzing text content and following hyperlinks. This information about the website is used to determine how it should be categorized and ranked. Because the spider's functions are so critical, anything relating to the way they operate is a closely guarded secret that the search engines would prefer you not to know. It is, however, in our best interest to understand as much as we can about them, and to use that information to our advantage when designing a webpage.

One of the most important things to keep in mind when designing your website is to see your site from a spider's point of view. A spider can only analyze text and words that are in a structured format. That is exactly why a frames-based site rarely ranks well on a search engine. The HTML for a frames site doesn't have a conventional format-- all of the content is jumbled about the page in different code sections and script excerpts, and that confuses spiders. Also, a spider needs to know right away what it should look for when it crawls your site. Using meta keywords is the best method for doing that. Otherwise, spiders will try to guess the content on your page and won't necessarily be successful -- getting ranked high for something unrelated to your site isn't helpful at all.

Descriptive and targeted meta keywords aren't the only thing search engine spiders look for. If you do use meta keywords, a spider tries to find out how relevant those keywords are for your site. For example, if your site is about recreational fishing, and you use the words fly-fishing, angling, and deep-sea fishing multiple times in your site, the spider will see your site being more relevant to those particular words than words which only appear once (for example, "commercial fishing"). Also, some spiders consider the position of a keyword to be important. If a keyword is in the page title, or in the first six lines of the page body, some search engine spiders consider that to be very significant. The "weight" of a keyword is a big factor, as well. If a keyword appears three times in a page with one thousand words, that keyword has a lower weight than if it was on a page with thirty words. Pages with heavily weighted keywords are considered more relevant to that keyword, and usually rank higher. However, it is possible to go too far and actually abuse the way a spider works.

While it is good to optimize your page, overdoing it can cause the spider to think that you are trying to fool it or spam the engine. The most common way of doing this is by using too many meta keywords. In an effort to rank their site higher, some webmasters will have an absurd amount of keywords. They'll include a meta keyword section two or three times in their page. Not only is this not effective, it is counter-productive. Something just as common is repeating a keyword over and over again on the page. Years ago it was useful to do this, but search engine spiders have advanced enough that simple tricks aren't going to fool them. Another, more devious plot, is called ghosting. When a spider accesses a site, it tells that site who it is. So, a webmaster can detect that a search engine spider is going to look at its site. Instead of serving up the normal webpage that is seen in a web browser, the webmaster gives the search engine spider a specially optimized page designed to rank perfectly on that engine. While this practice may seem good for pages with a lot of dynamic content and not a lot of text, it is still abusing the purpose of spiders. The webmasters who practice ghosting aren't only misleading the search engines, but they are also misleading web surfers coming to their site expecting to find the information they are looking for, but instead find themselves at a site which they didn't want to visit. The people behind search engines are always updating their spiders, making them both more effective and better able to sniff out sneaky webmasters trying to abuse the system.

Spiders are the workers behind the scenes at the search engines. Some of them crawl through millions of websites every month. A website's success depends on cooperating with the search engines and their methods. To cooperate with the search engines, it is also important to understand how they and their spiders operate. AddWeb 4's Page Advisor is designed to help you determine what spiders are looking for when they award a page a high ranking position. Also, by comparing your website to the top three sites on an engine, you can learn what other webmasters do that give them a high position, and you can incorporate those features into your site.

AddWeb Article- Copyright (c) 2000
Cyberspace Headquarters, LLC. All rights reserved.